



LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

Wind Power Curve Modeling in Complex Terrain Using Statistical Models

V. Bulaevskaya, S. Wharton, W. Miller

December 17, 2013

Journal of Renewable and Sustainable Energy

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

Wind Power Curve Modeling in Complex Terrain Using Statistical Models

V. Bulaevskaya,^{1, a)} S. Wharton,¹ A. Clifton,² G. Qualley,^{3, b)} and W. O. Miller⁴

¹⁾Lawrence Livermore National Laboratory, Livermore, CA 94551

²⁾National Renewable Energy Laboratory, Golden, CO 80401

³⁾Infigen Energy, Dallas, TX 75206

⁴⁾Lawrence Livermore National Laboratory, Livermore CA 94551

(Dated: 22 December 2014)

The simplest power curves model wind power only as a function of the wind speed at the turbine hub height. While the latter is an essential predictor of power output, wind speed information in other parts of the vertical profile, as well as additional atmospheric variables, are also important determinants of power. The goal of this work was to determine the gain in predictive ability afforded by adding wind speed information at other heights, as well as other atmospheric variables, to the power prediction model. Using data from a wind farm with a moderately complex terrain in the Altamont Pass region in California, we trained three statistical models—a neural network, a random forest and a Gaussian process model—to predict power output from various sets of aforementioned predictors. The comparison of these predictions to the observed power data revealed that considerable improvements in prediction accuracy can be achieved both through the addition of predictors other than the hub-height wind speed and the use of statistical models. To our knowledge, the use of the Gaussian process model in this context is new, in contrast to neural networks and random forests. The advantage of this model over the other two models is that it provides a much more natural way to estimate the uncertainty associated with its predictions. In addition, this study presents one of the most comprehensive analyses to date of the relative importance of various wind power curve inputs.

Keywords: wind power; power curve modeling; statistical modeling; machine learning; lidar; complex terrain

I. INTRODUCTION

As the United States and other countries set increasingly ambitious wind energy production goals, an ability to produce reliable wind power forecasts is becoming more and more critical¹. To be valuable, a forecasting tool must not only produce accurate point forecasts of power, but also correctly quantify the uncertainty associated with these point predictions. Statistical analysis and modeling are essential for designing a forecasting system that has both of these features.

Wind power forecasting involves converting atmospheric forecasts into a forecast of power output from an individual turbine or a set of turbines (e.g., all the turbines in a wind park). Even if a perfect atmospheric forecast were available, the model relating it to wind power is far from obvious. The most basic power curves model turbine power output as a cubic function of the wind speed at turbine hub height, assuming standard values of turbulence intensity and adjusting for air density in a simple fashion. This is a great oversimplification of reality. Although hub-height wind speed is essential, other variables, such as wind speeds in other parts of the vertical profile, atmospheric stability and wind veer can also be important determinants of wind power output².

Our aim is to develop more accurate power curves through the use of statistical modeling. In particular, we

model power production of a wake-free leading-row turbine in the hilly Altamont Pass region of California. We consider various sets of inputs beyond hub-height wind speed, including wind speed information in the entire profile of the rotor disk and wind veer. We also compare adding air density explicitly in the model to the standard industry practice of applying an air density adjustment to the wind speed. We evaluate three statistical models as approximations for the power curve, or the function relating these inputs to wind power output: neural networks (NN), random forests (RF) and Gaussian process models (GPM).

Besides potentially improving power forecast accuracy, statistical models have an additional advantage of quantifying the uncertainty associated with the forecast. This is in stark contrast to the deterministic nature of more traditional power curves as those only provide point forecasts. As will be discussed in Section V, the GPM, which has not been previously explored in the context of wind forecasting, is particularly well suited for this purpose. The capability to quantify power forecast uncertainty can be of tremendous value: for example, utilities companies can greatly benefit from this feature when they determine unit commitment schedules^{3–5}.

II. LITERATURE REVIEW

The simplest way to predict power generation of a specific turbine is to use a turbine manufacturer-supplied power curve (MPC). MPCs usually model power output as a function of mean hub-height wind speed, averaged over a 10-minute interval and adjusted for non-standard

^{a)}Corresponding author; Address: P.O. Box 808, L-211, LLNL, Livermore, CA 94551; Electronic mail: verab@llnl.gov

^{b)}Current address: Pentalum, Colleyville, TX 76034

air density and sometimes for a range of turbulence intensities. In complex terrain, these curves can become especially inaccurate because a significant portion of the rotor-disk flow is not horizontal and steady, as assumed by the curves, but instead includes upflow or downflow, as well as significant turbulence⁶⁻⁸. At the same time, in many parts of the globe, wind farms are increasingly being built in areas of complex terrain to take advantage of accelerated flows along ridgelines and on the crest of hills.

Studies have shown that variation in atmospheric conditions, including turbulence intensity, wind shear and wind veer, can lead to changes in power output of 10% or more for a given hub-height wind speed^{2,9-16}, thus making a strong case for supplementing hub-height wind speed with these variables in the model for power. We thus pursue two avenues for improvement over the basic power curves: first, supplementing the hub-height wind speed with other atmospheric variables as inputs to the power curve and second, using statistical models to approximate the power curve.

Our work builds upon a growing body of literature exploring one or both of these directions. One of the most recent examples is the use of an RF model to predict wind power output of a 1.5-megawatt (MW) turbine¹⁶. The authors simulated both the wind field inputs and the corresponding power output and compared predictions from an RF model to those of a traditional power curve method discussed earlier. They found that adding wind shear and turbulence intensity as predictors to the 10-minute average hub-height and using an RF resulted in up to a three-fold reduction in prediction error relative to the traditional power curve method. Similarly, in Ref. 17, the authors used simulated data to compare RFs to the method of binning and a physics-based model (one suggested in a 2013 draft revision of IEC 61400-12-1) and found RFs to be the most accurate.

RFs are also explored in Ref. 18, where they are compared to a number of other statistical models, including neural networks (discussed below), support vector machines and linear regression, as well as to the persistence model, in terms of their performance in predicting hourly power averages at time horizons ranging from 1 to 60 hours, at three different wind parks in France. The authors used forecast wind speed at 50 meters above ground and wind direction as input variables and found that the RF tended to have the best performance of all the models, yielding up to a three-fold reduction in normalized root-mean-squared error relative to the persistence model. Because this work and the studies in Refs. 16 and 17 suggest that RFs are a promising forecasting tool, we explore them in our work, as well.

A statistical model that is much more commonly encountered in wind power forecasting literature than an RF is the NN¹⁹. In some cases, such as in Ref. 20, NNs are first used to forecast wind speed, and this forecast is then converted to a power forecast using the traditional MPCs. Alternatively, NNs are employed to model power

output directly. For example, in Ref. 21 the authors describe using an NN to learn the relationship between atmospheric variables, such as wind speed, air pressure and temperature, and the power output, both for long- and short-term forecasting. In Ref. 22, an NN is developed to produce short-term forecasts of power at a given site using wind speed at the site and neighboring locations. The authors found up to 24% error reduction compared to the persistence model. In Ref. 23, power predictions at an individual turbine using an NN with wind speed and direction as inputs were associated with up to a ten-fold reduction in percent error relative to the traditional model. We explore NNs in this work, both because of their prominence and predictive skill advantage relative to traditional models.

In Ref. 24, on the other hand, the authors use conditional kernel density (CKD) estimation to forecast wind power given the distribution of wind speed forecasts. They first use time series modeling to produce forecast distributions of the hub-height wind speed and then CKD to model the uncertainty in the power curve given the wind speed forecast. By handling both of these sources of uncertainty, they establish a powerful modeling framework. We did not consider this type of an approach in our work, however, because density estimation becomes increasingly challenging as the number of variables grows, and we wanted to explore a much greater set of inputs to the power curve than the hub-height wind speed.

Time series modeling has also been frequently used to forecast wind power directly, with Refs. 25–27 being just a few of the examples. Furthermore, Ref. 28 gives a comprehensive survey of space-time modeling in wind power forecasting, which incorporates information at nearby locations in addition to power observed at the location of interest in recent past. In Ref. 29, a Markov chain model is used to predict the change in power 1 hour ahead. Although these types of models can be quite successful, they are typically only useful for very short-term forecasting (up to only a few hours). Moreover, our immediate aim in this work is not to develop an operational power forecasting model, but rather to gain a better understanding of the relationship between power and its key atmospheric determinants. Thus, we did not consider time series or space-time models in this work.

Compared to the studies mentioned above, in our work we consider a wider range of input variables (discussed in Section IV) and an additional class of models, GPMs (discussed in Section V), which we have not encountered in wind power forecasting literature so far.

III. DATA

In June 2012, a vertically-profiling, ground-based Doppler Light Detection and Ranging (lidar) instrument (Wind Cube v2, Leosphere, Orsay France) was deployed at a wind farm in the Altamont Hills region of California for a 2-month field campaign. The wind farm consists of

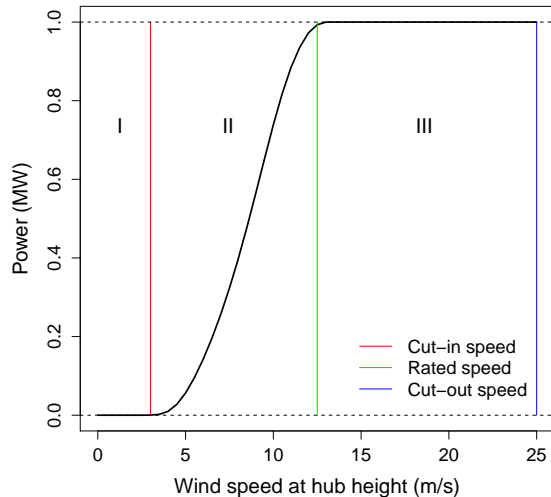


FIG. 1. The MPC (power in MW vs. hub-height wind speed in m/s) for the MWT62-1000A turbine used in the study³⁰, with the vertical lines indicating the cut-in, rated and cut-out speeds listed in Table I and Roman numerals labeling the three regions of the power curve defined by these speeds.

38 1-MW Mitsubishi (MWT62-1000A) blade-pitch controlled turbines. The turbine specifications are listed in Table I, and the corresponding MPC is shown in Fig. 1.

TABLE I. Specifications of the MWT62-1000A turbines used in the study³⁰.

Wind class	IEC Class IIA	Hub height	55 m
Cut-in speed	3 m/s	Rotor diameter	61.4 m
Rated speed	12.5 m/s	Blade length	29.5 m
Cut-out speed	25 m/s	Rated power	1 MW

The lidar collected continuous wind flow measurements from June 7 to August 24, 2012, resulting in nearly 2,000 hours of archived data. The summer period was chosen because it is the peak wind season for this area due to strong sea breezes. The lidar provided measurements of wind flow in the streamwise, crosswind, and vertical directions at 12 measurement heights from 40 to 150 meters above ground level (a.g.l.) in steps of 10 meters with a probe volume of 20 meters and sampling frequency of 1 Hz. The data were averaged over 10-minute intervals to calculate the mean and standard deviation of the horizontal and vertical wind speeds, as well as wind direction at each sampling height. The data were also quality-controlled for signal-to-noise ratio and availability: data points with carrier-to-noise ratio (CNR) below -23 and availability less than 25% during each 10-minute averaging period were discarded. Wind speed uncertainty is estimated to be 3–5% for moderately complex terrain sites³¹.

The Supervisory Control and Data Acquisition

(SCADA) system provided 10-minute averages of ground-level air density, hub-height wind speed and power output for all turbines, for the same 10-minute intervals as the lidar data. The SCADA data also contained other statistical summaries, including standard deviations, for these 10-minute intervals. Wind speed was measured with a nacelle-mounted cup anemometer (NRG #40, NRG Systems, Inc., Vermont, USA) at a frequency of 1 Hz.

The landscape surrounding the wind farm is moderately complex. The predominant wind direction is from the southwest during the summer, as shown in Fig. 2. The terrain’s influence on the wind patterns at the site is described in detail in Ref. 32. Because the turbines are located on a ridgeline with a slope of 10 degrees, the lidar had to be placed at a lower elevation to capture inflow conditions at a distance close to 2–4 rotor diameters (D), as recommended by the International Electrotechnical Commission (IEC)³³. Of all the heights measured by the lidar, the 10-minute wind speed averages of lidar measurements at 90 meters had the highest correlation and the smallest root-mean-squared error (RMSE) relative to the corresponding nacelle wind speed averages (correlation = 0.94, RMSE = 0.84 m/s). Fig. 3 shows the distribution of the differences between the nacelle and the 90-meter lidar wind speed means. The nacelle wind speed exceeds the lidar measurement by 0.4 m/s on average, indicating the frequent occurrence of accelerated flows at the top of the ridge. These flows, known as “hill speed-ups,” are discussed in detail in Ref. 32.

The examination of the wind rose at 90 meters (shown in Fig. 2) revealed that of the downwind turbines, turbine #8 (T8) had the highest frequency of being directly downwind of the lidar. Further analysis was therefore limited to this turbine. Fig. 4 shows the setup of the lidar and T8 turbine. As Fig. 2 shows, wind directions in the range 225°–250° were ideal for the lidar to capture inflow conditions seen by T8, so the data were restricted to these wind directions. The data were further limited to the time periods without any missing SCADA data.

In this study, we focused only on the 90-meter wind speeds that correspond to region II, or the portion of the power curve where power is most sensitive to small changes in wind speed. For an MPC, this corresponds to the range between the cut-in and the rated hub-height wind speeds, i.e., between 3 m/s and 12.5 m/s in our study, as can be seen from Table I and Fig. 1. Since our data are 10-minute averages, we adjusted this range so that the observed 10-minute power averages are largely constant (very close to 0 or rated power) outside the chosen limits and are thus minimally contaminated by observations from regions I and III. This adjustment led us to restrict the data to 90-meter average speeds between 3.5 m/s and 11.5 m/s (see Fig. 5). This and all the other filters discussed above resulted in 1737 data points, each representing a 10-minute period of lidar wind flow and power data.

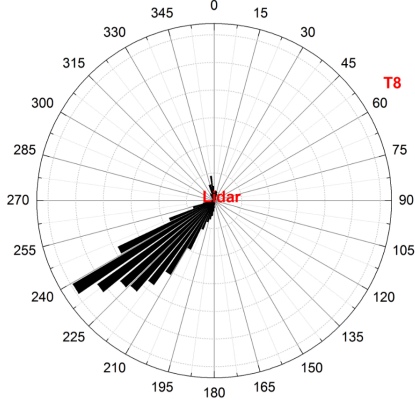


FIG. 2. Wind rose showing the distribution of wind direction (degrees) at 90 meters above ground level for the campaign period, as measured by the lidar. The figure also shows the location of T8 turbine in relation to the lidar (in red).

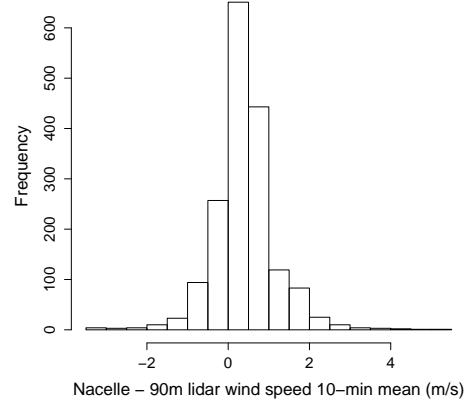


FIG. 3. The distribution of the differences between the nacelle and the 90-meter lidar wind speed means (m/s), with a positive difference implying that the nacelle mean is the larger of the two (RMSE = 0.84 m/s, correlation = 0.94).

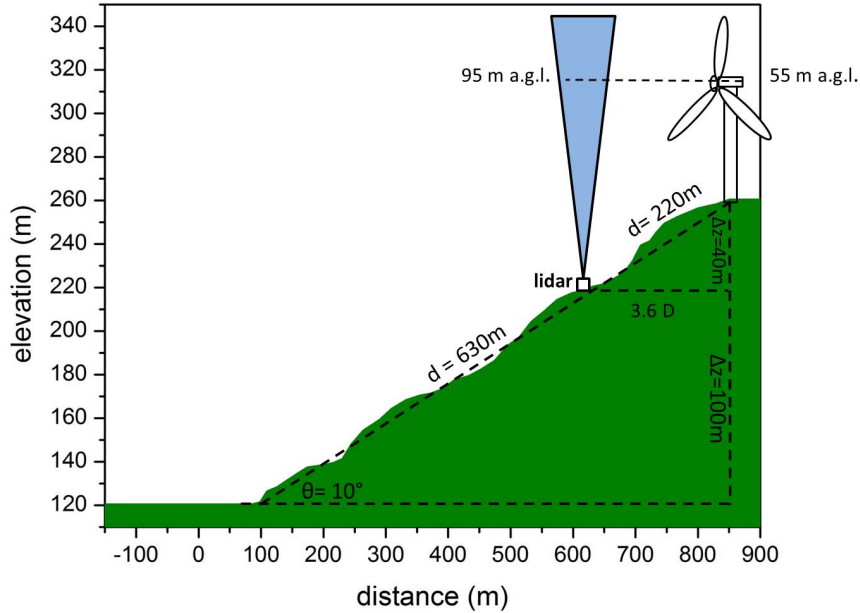


FIG. 4. The setup of the lidar and the T8 turbine relative to one another. The lidar was located 3.6 rotor diameters (D) upwind of T8.

IV. INPUTS TO THE POWER CURVE MODEL

A. Traditional power curve inputs

Traditional power curves model the expected 10-minute average of power for a given 10-minute average of the hub-height wind speed. These curves are typically

generated assuming the sea-level air density (or standard air density) of 1.225 kg/m^3 . However, in areas of high elevation or exposure to large air temperature ranges, air density may differ substantially from this value at any given time. To account for this potential discrepancy between the observed and the standard air densities ρ and ρ_0 , the IEC recommends adjusting the 10-minute average of the hub-height wind speed U for the observed air

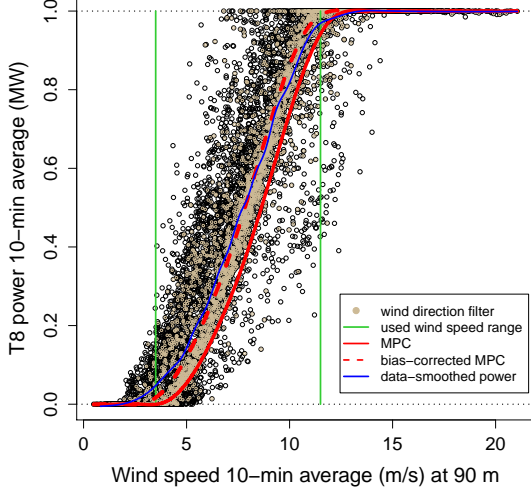


FIG. 5. Ten-minute average power (MW) at turbine T8 vs. 10-minute average wind speed (m/s) measured by the lidar at 90 meters above ground (black circles). Also shown are the data filters described in Section III, the predictions using the MPC and the bias-corrected MPC (see Section VII), and the data-smoothed power curve as a function of the wind speed mean at 90 m (see legend).

density ρ and plugging this density-adjusted value $U_{d.adj}$ into the power curve model.

The recommended adjustment is given by

$$U_{d.adj} = U \left(\frac{\rho}{\rho_0} \right)^{1/3} \quad (1)$$

and is based on the following first principle-based model for power output P at a given time point:

$$P = AC_p \rho U^3, \quad (2)$$

where A is the area of the turbine's rotor disk, C_p is the power coefficient at that time point, and ρ and U are as above³⁴. Rewriting Eq. (2) in terms of ρ_0 yields

$$P = AC_p \rho_0 \left[U \left(\frac{\rho}{\rho_0} \right)^{1/3} \right]^3 = AC_p \rho_0 U_{d.adj}^3, \quad (3)$$

resulting in the adjustment in Eq. (1). Note that applying this adjustment to 10-minute averages of hub-height wind speeds implicitly assumes that air density tends to be constant over a 10-minute period. As discussed in Section II, the model in Eqs. (2) and (3) is an oversimplification of reality, both because it uses a very limited set of inputs and because it assumes a very particular relationship between these inputs and power.

B. Expanded set of power curve inputs

Our goal is to determine the gain in predictive ability afforded by adding information other than the density-adjusted hub-height wind speed to the model, as well as by assuming less constrained forms of the relationship between these inputs and power. In particular, the data described in Section III allowed us to examine the value of including wind speed means and standard deviations at heights other than the hub height, as well as adding wind veer to the model. In addition, we explored using air density explicitly in the model, by including it in addition to the *unadjusted* wind speed, instead of using the density adjustment in Eq. (1) prescribed by the model in Eqs. (2) and (3).

These additional variables were considered because, as discussed in Section II, they or quantities very closely related to them have been shown through various studies to be among the most important atmospheric factors for wind power production after the hub-height wind speed. For example, wind shear is captured by including the wind speed profile and wind veer in the model. Similarly, the information conveyed by turbulence intensity is represented by including the means and standard deviations at the hub height and other heights of the vertical profile.

The studies cited in Section II also identified wind direction, temperature and pressure as important inputs to the power forecast model. Although we did not use wind direction directly as an input in our study, it was accounted for indirectly by using only the data corresponding to the range of directions favorable to the turbine to build our models, as explained in Section III. Likewise, we did not consider temperature and pressure because by including the air density, we largely capture the information in these two quantities.

1. Wind speed inputs

Although the SCADA data provided wind speed measurements taken at the nacelle of the turbine, these measurements can deviate from the true inflow conditions because the cup anemometer is subject to flow distortion from the blades and the nacelle hub. Consequently, we considered lidar measurements of free-stream wind speed rather than the nacelle measurements as the baseline in this study. The appropriateness of using the former as the baseline is discussed in more detail in Section VII.

Recall from Section III that lidar wind speed averages measured at 90 meters were in highest agreement with the wind speed averages observed at the nacelle. As a result, measurements at this height were considered a proxy for free-stream wind speed. The distribution of these measurements is shown in Fig. 6(a).

First, we modeled the power output at T8 as a function of density-adjusted wind speed variables (using the adjustment in Eq. (1)) alone. One way to include infor-

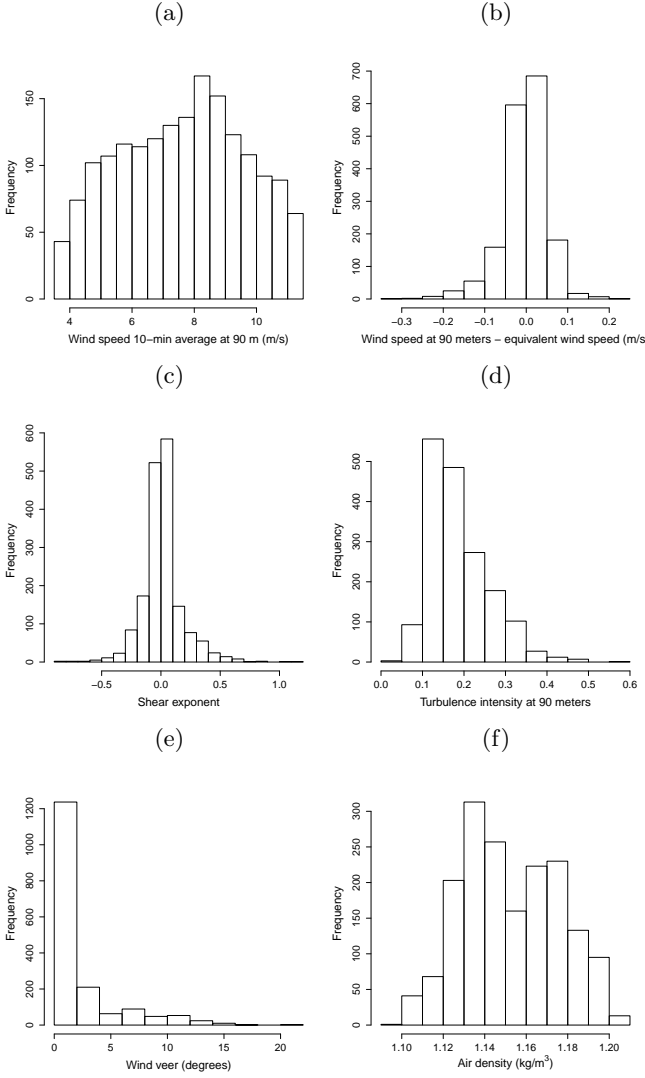


FIG. 6. The observed frequency distributions of the following (based on 10-minute means): (a) 90-m wind speed (m/s), (b) difference between the 90-m wind speed and the REWS, defined in Eq. (4) (m/s), (c) wind shear exponent α , (d) TI at 90 m, (e) wind veer (degrees), as defined in Section IV B 2, and (f) ground-level air densities (kg/m^3).

mation about the entire vertical profile of wind speeds is to estimate the average inflow seen by the entire rotor disk of the turbine. This can be done using the rotor-equivalent wind speed (REWS), defined in Ref. 35 as

$$\text{REWS} = \frac{2}{A} \int_{H-r}^{H+r} U(z)(r^2 - H^2 + 2Hz - z^2)^{1/2} dz, \quad (4)$$

where A is the area of the rotor disk, H is the turbine hub height, r is the rotor disk radius, $U(z)$ is the mean wind speed at height z , and dz is the height interval at which the wind speed measurements are taken. Note that this is just the weighted average of the mean wind speeds in each height interval across the rotor disk, where the weight for each interval is equal to the area occupied by the

interval as the fraction of the total area of the rotor disk. Since we used lidar measurements at 90 m as a proxy for free-stream wind speed, and the rotor disk radius for T8 is 30.7 m, the REWS in our case corresponded to the weighted average of lidar measurements from 60 to 120 meters above ground.

Fig. 6(b) shows the distribution of the differences between the 10-minute average wind speed at 90 meters and the REWS. It reveals that the magnitude of these differences tends to be small: for 74% of the observations the absolute difference is within 0.05 m/s. This is because, as can be seen from Fig. 6(c), the magnitude of wind shear is frequently very low at this site. Wind shear is defined as the value of α in the power-law profile of wind speed $U(z) = \beta z^\alpha$.

In addition to collapsing the vertical profile of lidar wind speed means into a scalar quantity such as REWS, we also explored the value of including the entire profile in the model directly. Moreover, we studied the effect of supplementing this profile with the profile of wind speed standard deviations, as these provide estimates of turbulence in the mean wind flow.

Turbulence can contribute valuable information about wind power production. On the one hand, it can enhance power production since turbines extract power from turbulence, as well as the mean wind flow, as seen in Ref. 2. On the other hand, its impact on power production can be negative because strong, intermittent bursts of turbulence, including coherent structures produced underneath a nocturnal low-level jet, can put high fatigue loads on the turbine's components and alter turbine response³⁶. For reference, the distribution of one of the most commonly used measures of turbulence, the turbulence intensity (TI), is shown in Fig. 6(d). It is defined as the ratio of the standard deviation to the mean of wind speeds in a 10-minute period at a given height (90 meters in the plot). Note that the MPC in Fig. 5 was generated assuming the TI of 10%, whereas Fig. 6(d) shows that for an overwhelming majority of observations (94%), it exceeds this value. High TI values are common in moderately complex terrain.

As was previously done with the 90-meter wind speed mean, we applied the density adjustment in Eq. (1) to the entire vertical profile of means. In addition, following the aforementioned assumption that air density is constant over a 10-minute period, the 10-minute standard deviations were adjusted in a similar manner:

$$SD_{d.adj}(U) = SD(U) \left(\frac{\rho}{1.225} \right)^{1/3}. \quad (5)$$

Consequently, we explored the following five sets of wind speed variables, all adjusted for air density and henceforth collectively referred to as Group 1 (the parentheses after each input set contain the label used for that set in the plots in Section VII):

1. Wind speed mean at 90 meters (*ws 90m mean*),
2. Wind speed mean at 90 meters + standard deviation at 90 meters (*ws 90m mean + sd*),

3. Rotor-equivalent wind speed, defined in Eq. (4) (*ws equiv*),
4. Entire profile of wind speed means at the 12 heights (40, 50, ..., 150 meters) (*ws all mean*) and
5. Entire profile of wind speed means + entire profile of wind speed standard deviations at the 12 heights (*ws all mean + sd*).

Note that the first set in the above list corresponds to our baseline method of modeling power as a function of the density-adjusted wind speed at the hub height alone. The other sets, on the other hand, include additional information, either in the form of the temporal variability in the wind speed (set 2) or wind speeds measured in other parts of the profile (sets 3 and 4), or both (set 5).

Furthermore, because the dimensions of the last two sets above are quite large (12 and 24 for sets 4 and 5, respectively) and the wind speeds at various heights are highly correlated (correlation values range between 0.91 and 0.98), these two sets were also subjected to a principal component analysis (PCA, see Ref. 37) in order to reduce their dimensions, thereby greatly simplifying fitting of the statistical models discussed in Section V. The top components that together accounted for at least 99% of the total variance were retained. Note that the aforementioned Group 1 of input sets includes the PCA-reduced versions of sets 4 and 5.

2. Additional atmospheric inputs

The data revealed that 10-minute wind direction averages can be quite variable across the vertical profile, motivating the inclusion of wind veer as an input. A large amount of wind veer in these data is not surprising because in complex terrain, local hills and valleys can create distinct flow features, especially at heights closer to the ground. We used the circular standard deviation (see Ref. 38) of all wind direction averages measured by the lidar in the 40–90 meter height range as a way to quantify wind veer. This range was used because wind direction averages at these heights had much higher circular correlations (see Ref. 38) with the nacelle wind direction average than those at the other heights. Fig. 6(e) shows the frequency distribution of wind veer defined in this way. As can be seen from it, although wind veer tended to be small most of the time, it was sizable for a non-negligible fraction of observations. Hence, to account for the potential effect of wind veer on power, we added it to each of the input sets in Group 1, thus creating another group of input sets, referred to as Group 2.

In addition to the wind flow across the rotor disk, the density of air is the other key factor affecting the amount of kinetic energy available for extraction from a turbine. To account for the effect of air density on power, in addition to using the standard air density adjustment in Eq. (1), we also considered the approach of adding the

10-minute air density average at ground level as a separate input to the *unadjusted* versions of each of the input sets in Group 1, thus creating Group 3. The frequency distribution of air density values, shown in Fig. 6(f), reveals that all observed densities were smaller than the standard value of 1.225 kg/m^3 , with a large fraction being considerably smaller than this value as a result of the wind farm’s elevation. Thus, air density is expected to be an important factor for determining power at this location.

Finally, Group 4 of input sets was created by adding both wind veer and air density to the *unadjusted* wind speed input sets in Group 1. The four groups thus resulted in a total $4 \times 5 = 20$ input sets. The variables included in each group are summarized in Table II.

TABLE II. Summary of the power model input variable combinations considered in our analysis. An “x” under a variable name indicates its inclusion in the group. Each group consists of 5 sets, made up of the wind speed variables enumerated in Section IV B 1 (referred to collectively as “Wind speed” in the table) and all of the variables marked with an “x”. Note that in Groups 1 and 2 the air density adjustments in Eqs. (1) and (5) are applied to the wind speed variables, while no adjustment is applied to these variables in Groups 3 and 4.

Group	Input variables		
	Wind speed	Wind veer	Air density
1	x		
2	x	x	
3	x		x
4	x	x	x

V. STATISTICAL MODELS

Basic power curve models cannot accommodate input sets that include variables other than the density-adjusted hub-height wind speed, motivating the use of statistical models to predict power for the more complex input sets. Moreover, even in the absence of the additional input variables, a statistical model has two advantages alluded to earlier. One advantage is that it does not assume a particular relationship between the inputs and the output, as these relationships can be learned by the model from the data. Another is that statistical models provide an estimate of the uncertainty associated with the predictions.

We compared three statistical models in terms of their ability to predict power output using each of the input sets described in Section IV. These were NNs, RFs and GPMs. While the reader is referred to Refs. 39 and 40, Refs. 40 and 41, and Refs. 42–44 for a detailed treatment of NN, RF and GPM, respectively, Appendix A gives basic details of each model.

The statistical software R⁴⁵ was used to train all three models. Each of the models requires the user to specify certain parameters. As discussed in more detail in

Appendix A, for some of these parameters we used well-established default values or functional forms, while others were determined using cross-validation or maximum likelihood estimation. In addition, as described in more detail in Appendix A3, we used two different forms of the mean function for the GPM: a constant and a natural spline. These will be referred to in the remainder of the paper as GPM-c and GPM-ns, respectively.

While NN and RF can be very powerful prediction tools (as discussed in Section II), the GPM provides a much more natural framework for modeling prediction uncertainty. This was the primary motivation for including the GPM in the analysis because, as mentioned in Section I, the capability to produce accurate uncertainty estimates associated with point predictions is a highly desirable feature in a wind power forecasting tool.

The GPM models the joint distribution of all the observed and predicted response values as a multivariate Gaussian random variable, providing a well-established probabilistic framework for the entire distribution of predictions. This leads to straightforward inference about forecast uncertainty, quantiles and tail probabilities, and at a modest additional computational cost relative to that required to get the point predictions themselves. In contrast, NN and RF model the response quite differently from the GPM: being regression models, they treat each response value as conditionally independent of the other response values given the parameters. In addition, the conditional distribution of the prediction given the parameters is not easily tractable for NN and RF (unlike linear regression, for example). Consequently, there is no theoretical foundation for exact probabilistic inference for NN and RF predictions, and one must resort to approximations-based on bootstrap (as in the case of NN) or an empirical distribution of predictions in the ensemble (as in the case of RF)—to provide estimates of prediction uncertainty.

At the same time, we must also note a disadvantage of GPM over the other two models: it is not as scalable as NN and RF in terms of the number of predictors it can accommodate easily. If the predictor set is large, training GPM is more computationally expensive than NN and RF and becomes infeasible at smaller thresholds than the other two models. However, since training of the model can be done offline, as long as it is feasible, the time required to train the model does not pose a great concern (for reference, the largest input set in this study consisted of 27 predictors, and none of the sets encountered the infeasibility issue). In addition, as discussed in Section IV, dimension reduction tools, such as PCA, can be used to make the size of the predictor set more manageable for the GPM.

There was evidence of slight heteroscedasticity, or non-constant variance of power as a function of the inputs, in the data. Consequently, we also explored a variant of the Gaussian process model, a treed Gaussian process model (TGP), because it is able to identify different variance regions in the data and model them separately, thus ac-

counting for heteroscedasticity^{46,47}. However, the TGP results were practically the same as those for GPM, and since the computational burden of TGP is considerably higher than that of GPM, it was not considered further.

VI. PERFORMANCE STUDY

To assess the performance of each of the sets of inputs and the three statistical models outlined above, the 1737 observations consisting of the lidar and SCADA data were randomly divided into two mutually exclusive sets of 869 and 868 points, the first of which was used to train each of the models with each of the input sets in each group. For comparison, using the same training set, we also trained each of the statistical models with the 10-minute mean and standard deviation of the nacelle wind speed as inputs.

Each combination of the trained model and input set was then used to make wind power predictions for the remaining 868 points in the data, and these predictions were compared to the observed wind power values for these points (henceforth referred to as the validation set). In addition, we obtained predictions of power using the MPC³⁰ (shown in Fig. 1) using the same training and validation sets.

This experiment of partitioning the data, training the models with the training set and making predictions for the validation set was repeated a total of 30 times to obtain a representative set of results. For each experiment and model/input set combination, the RMSE of 10-minute average wind power predictions relative to the actual SCADA data was computed as a metric of predictive skill. The RMSE for the i^{th} experiment is given by

$$RMSE_i = \sqrt{\frac{\sum_{j=1}^{868} (P_{ij} - \hat{P}_{ij})^2}{868}}, \quad (6)$$

where P_{ij} and \hat{P}_{ij} are the actual and predicted power value, respectively, for the j^{th} observation in i^{th} experiment's validation set with $i = 1, \dots, 30$ and $j = 1, \dots, 868$.

VII. RESULTS AND DISCUSSION

A. Wind speed inputs

We first consider the results for the input sets involving only the wind speed variables (Group 1). Fig. 7(a) shows the RMSEs for the 30 experiments described in Section VI using lidar measurements of the wind speed as inputs to each of the statistical models. Also shown in the plot are the RMSEs for the predictions obtained using the bias-corrected MPC using the density-adjusted wind speed mean at 90 meters as an input. The bias correction is motivated by the fact that the original MPC

tends to underpredict power, as can be seen from the solid red curve in Fig. 5. Such a consistent bias can be easily corrected by shifting the curve horizontally, to the left in this case (in the case of consistent overprediction, the necessary shift would be to the right). The value for the shift can be chosen by finding the value that minimizes the RMSE (or any other sensible measure of discrepancy) of the predictions relative to the data.

We apply this correction in our study since we assume that any user of an MPC would do so given a sufficient amount of data, and we treat the bias-corrected MPC as the baseline. In addition, the fact that we use the lidar—rather than the nacelle—wind speed data makes the bias correction even more critical for a fair comparison between the traditional and our power curve models. The bias-corrected MPC is shown in Fig. 5 (red dashed curve) and is close to the data-smoothed curve (in blue), indicating that the correction is reasonable.

The plot in Fig. 7(a) shows that among the five input sets we considered, the set with just the wind speed mean at 90 meters (the 1st cluster of boxplots) is one of the worst performers in terms of the RMSE, regardless of the model. Moreover, for this input set the bias-corrected MPC and the statistical models tend to perform equally in terms of prediction accuracy. This suggests that when only the 90-meter free-stream wind speed mean data are available, the main reason to prefer statistical models over the MPC is the ability of the former to produce uncertainty estimates associated with predictions.

The comparison of the first two clusters of boxplots in Fig. 7(a) reveals that adding the 10-minute standard deviation information at 90 meters results in an average RMSE reduction of 0.007–0.008 MW, depending on the model. Reductions of such magnitude are practically significant considering that they represent the error differences in just one given 10-minute period and thus can add up to a substantial difference over time. These differences are also statistically significant (see below) at a 0.1% significance level (indeed, the RMSE was lower as a result of this addition in *each* of the 30 experiments). The impact of adding the standard deviation of the wind speed to the model is not surprising because it reflects turbulence in the wind flow, which is an important factor in power production, as discussed in Section IV. Its importance can be particularly high when the inputs are limited only to the measurement at the level equivalent to the hub height.

To judge the statistical significance of these and other RMSE differences we report, we computed the bootstrap confidence interval for the average reduction in RMSE in a random experiment for each of the models. Specifically, we used the nonparametric bias-corrected accelerated percentile confidence interval (BC_a CI) method, discussed in detail in Ref. 48 (see Appendix B for a discussion on the applicability of the bootstrap methodology in this context). The *bootstrap* package in R⁴⁹ was used to calculate the intervals. Since we wanted to compare many pairs of input sets and models, to control for the

rate of falsely declaring a difference between any of these pairs, we used a very high confidence level of 99.99% for each comparison, which is equivalent to a 0.01% significance level (see Ref. 50).

The REWS (3rd cluster of boxplots in Fig. 7(a)) performs only slightly better than the wind speed mean at 90 meters. Although the average RMSE reduction is significantly above 0, it is quite modest, at less than 0.003 MW for all four models. Such a small difference between the REWS and the 90-meter wind speed mean is due to the fact that, as discussed in Section IV and shown in Fig. 6(b), the REWS differed very little from the wind speed at 90 meters for the majority of observations, which in turn is the consequence of generally low wind shear at the site (as shown in Fig. 6(c)).

Using the entire profile of wind speed means (4th cluster of boxplots in Fig. 7(a)), on the other hand, leads to a considerable and statistically significant reduction in RMSE relative to using the 90-meter wind speed mean alone, ranging from 0.02 to 0.04 MW on average (minimum reduction across the 30 experiments ranged from 0.006 to 0.03 MW, depending on the model). Adding the standard deviation profile to the mean profile (5th vs. 4th cluster of boxplots in Fig. 7(a)) results in very little benefit, however (the differences are either statistically or practically insignificant). This is most likely because the mean profile largely captures the turbulence information contained in the standard deviation profile.

Reducing the PCA-transformed profile of wind speed means by retaining the top components that together accounted for 99% of the variance led to using only 3 to 4 principal components (versus the original 12) in each of the experiments. Such a large dimension reduction was possible because, as mentioned in Section IV, the wind speeds at the 12 heights are highly correlated. As a result, as the comparison of every other boxplot to the preceding one in the last two clusters of boxplots in Fig. 7(a) reveals, the dimension-reduced wind speed profile performs almost as well as the original profile for each of the models. The mean RMSE difference between the reduced and the original predictor set is negligible, ranging from 0.003 to -0.001 MW across the models (with a negative value implying that the reduced set had the smaller RMSE, which was the case for RF).

The results are very similar when comparing the input set that also includes the entire profile of standard deviations to its PCA-reduced counterpart (in this case, the number of components decreased from the original 24 to 11 or 12). As discussed in Section V, reducing the dimension of the predictor set can greatly simplify the fitting of the models, particularly the GPM, so the fact that very little is sacrificed in terms of performance as a result of the reduction is highly valuable.

The statistical models tend to perform more or less equally (the highest average difference was 0.0003 MW over all pairs of models). The fact that GPM performs no worse than the other two models is important because, as discussed earlier, it has the advantage of providing

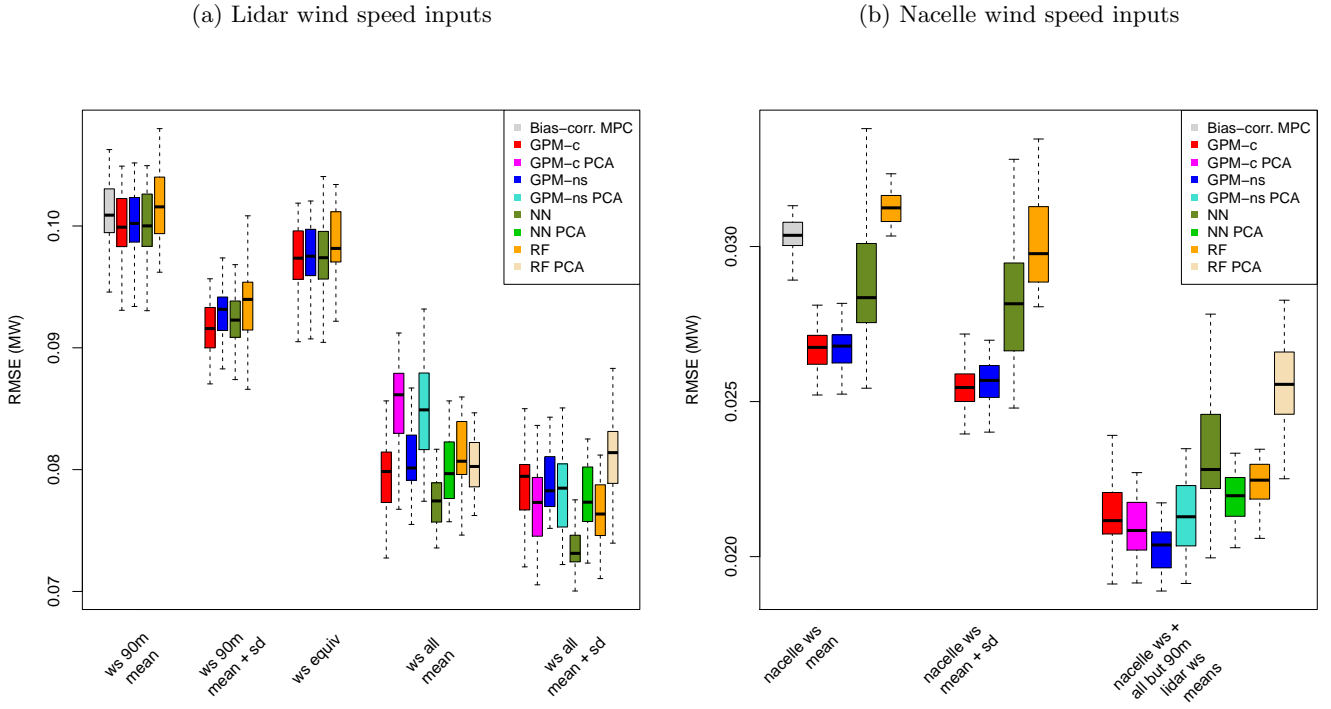


FIG. 7. Boxplots of RMSE values (MW) for the 30 experiments described in Section VI using (a) lidar wind speed inputs (sets 1–5 in the list in Section IV B 1) and (b) either nacelle only or a combination of nacelle and lidar wind speed inputs, as described in the text. The colors represent the statistical models or the MPC as indicated in the legend. The PCA-reduced versions of each model are denoted with “PCA” after the name of the model. Note that the scales on the y -axis are different in the two plots.

uncertainty estimates in a much more natural manner than NN and RF. Given this advantage and comparable performance to that of NN and RF, GPM may be the preferred model in this case.

We now consider nacelle measurements as inputs to the power model. The 1st cluster of boxplots in Fig. 7(b) shows the performance of the bias-corrected MPC and the statistical models using the nacelle wind speed mean as an input. The 2nd cluster of boxplots shows the same information, but for the input set that also includes the nacelle wind speed standard deviation. In contrast to lidar measurements, there is almost no benefit from adding the standard deviation of wind speeds to the input set (the mean RMSE difference is only 0.001 MW for each of the models). These findings are consistent with those in Ref. 2, in which the authors found no benefit from including nacelle-based TI in the measured power curves. This is because nacelle-mounted TI measurements are poorly correlated to free-stream stability⁵¹.

In addition, nacelle wind speed measurements are much better predictors of power on their own than those obtained with the lidar, as suggested by much lower RMSEs of the former compared to the latter (the mean difference in RMSE is 0.07 MW for the MPC and 0.06–0.09 MW for the statistical models, and all are statistically significant), so the benefit of adding standard deviation information is marginal. The superiority of nacelle mea-

surements relative to lidar ones is due to the fact that the inflow seen at the nacelle is not well reflected in the lidar measurements: recall from Section III that even though of all the heights the 90-meter wind speed means had the highest correlation and the smallest RMSE relative to the nacelle means, the RMSE was nevertheless considerable, at 0.84 m/s (Fig. 3 shows the distribution of the differences between these two measurements).

The discrepancy between the lidar and nacelle measurements is likely caused by the interactions between the flow and the complex terrain at the wind park (see Ref. 52 for an example of this phenomenon). We expect that a location with a simpler terrain will not experience such a high discrepancy between these two sources of measurements, and we are currently collecting data to test this hypothesis.

Although in our study the lidar did not provide a more accurate power prediction than the nacelle wind speed, there are several reasons for utilizing lidar in power curve studies. First, as mentioned earlier, in flat terrain an upwind lidar should provide accurate measurements of inflow conditions across the entire rotor disk, in which case it is expected to improve power prediction relative to using nacelle-based wind speed measurements at the hub height only. Second, even in complex terrain, supplementing the nacelle wind speed with lidar measurements at other heights can reduce prediction error. The

last cluster of boxplots in Fig. 7(b) shows the RMSE values for the input set consisting of the nacelle wind speed mean and the means measured by lidar in the entire vertical profile except for 90 meters (the latter was excluded because it is redundant with the nacelle measurement). Compared to the clusters of boxplots with just the nacelle inputs in the same plot, these tend to be significantly lower. The addition of lidar measurements to the nacelle measurement resulted in an RMSE reduction in each of the 30 experiments for all models, and the mean RMSE reduction ranged from 0.006 to 0.009 MW across the models. Third, lidar provides measurements of additional atmospheric variables, which can significantly improve power forecast accuracy, as will be discussed in more detail in Section VII B.

To illustrate wind power predictions and their associated uncertainty, we randomly chose one out of 30 experiments. Fig. 8(a) shows the predicted power (circles) and the associated 95% prediction intervals (shaded areas) for this experiment's validation set, obtained with the GPM-c model and using either the 90-meter wind speed mean alone (red) or the entire PCA-reduced profile of wind speed means (blue) as inputs. The plot shows prediction results for each of these two input sets, as well as the observed power for this validation set (grey triangles), as a function of the wind speed at 90 meters. The plot in Fig. 8(b) shows the errors in the predictions (i.e., predicted minus observed power, shown in Fig. 8(a)) with the same 95% prediction intervals as in Fig. 8(a), but shifted by the observed power, just like the predictions.

The point predictions that result from using the entire wind speed mean profile tend to be closer to the observed values than those using just the 90-meter wind speed mean, as indicated by their closer average proximity to the line of zero error, or perfect prediction (solid grey), in Fig. 8(b): the RMSE is 0.074 MW and 0.097 MW for the former and the latter, respectively, compared to 0.099 MW for the bias-corrected MPC predictions for this particular experiment. Moreover, both plots show that the uncertainty estimates corresponding to the PCA-transformed entire profile (blue) tend to be more narrow than those corresponding to the 90-meter wind speed mean alone (red). The average 95% prediction interval length for this validation set is 0.27 MW (corresponding to an uncertainty of ± 0.135 MW) and 0.36 MW (an uncertainty of ± 0.18 MW) when using the entire vertical profile and the 90-meter wind speed only, respectively. This confirms the earlier conclusion that using the entire profile can markedly improve prediction quality compared to using just the hub-height wind speed alone.

Recall our claim in Section IV B that wind shear information is captured by including the entire profile of wind speed means in the model, and that similarly, TI is represented by including both the means and standard deviations. Fig. 9 suggests that this is indeed the case. It shows the boxplots of RMSEs for the same experiments as described above, as well as either the shear

exponent α or the TI added to some of the wind input sets described in Section IV B 1. While adding shear to the 90-meter wind speed mean (2nd vs. 1st cluster of boxplots) leads to a small improvement in RMSEs for all statistical models, a much more pronounced reduction is achieved by using the entire profile of wind speed means (5th cluster of boxplots), including the PCA-reduced profile. This is not surprising since wind shear, which reflects the shape of the wind's vertical profile, only partially captures the information found in the entire profile of wind speed means. Moreover, as the plot shows, even using only the 90-meter wind speed mean and standard deviation (4th cluster of boxplots) leads to better results than using 90-meter wind speed and wind shear. As expected, adding shear to a full profile of wind speed means (6th vs. 5th cluster of boxplots) does not lead to an improvement in RMSE because the information captured in the former is already conveyed by the latter.

Adding TI at 90 meters to the 90-meter wind speed mean (3rd cluster of boxplots) leads to the same reduction in RMSE as adding the standard deviation at 90 meters (4th cluster of boxplots). However, when the entire profile of wind speed means is used, adding the entire profile of wind speed standard deviations results in a better performance for some of the statistical models than adding the entire profile of turbulence intensities (8th vs. 7th cluster of boxplots). Thus, allowing the model to use the raw means and standard deviations leads to the same or better results than transforming these to turbulence intensities.

B. Additional atmospheric variables

Here we examine the results for groups of variables that include inputs other than the wind speed information, specifically, wind veer, air density added separately to the unadjusted wind speeds and the combination of the two (Groups 2–4 outlined in Section IV B 2 and listed in Table II). GPM-c and NN tended to be the best-performing models in the results shown above and in all the other groups of variables, so in this section we limit our discussion only to these two models. Furthermore, the PCA-reduced versions of all the larger wind profiles (sets 4 and 5 in the enumerated list in Section IV B 1) tended to perform virtually the same as the original profiles for all the groups, so we only show the results for the former when discussing the larger sets.

Fig. 10(a) shows the percent reduction in RMSE due to adding wind veer to the model already containing density-adjusted wind speed inputs (Group 2 in Table II). This is the difference in RMSE for the models with and without wind veer as a percentage of the RMSE of the latter model, with a positive reduction implying higher RMSE for the latter model. For all input sets and models, adding wind veer resulted in an RMSE reduction in every one of 30 experiments. For the 3 smaller wind input sets, i.e., those consisting of either only the wind speed in-

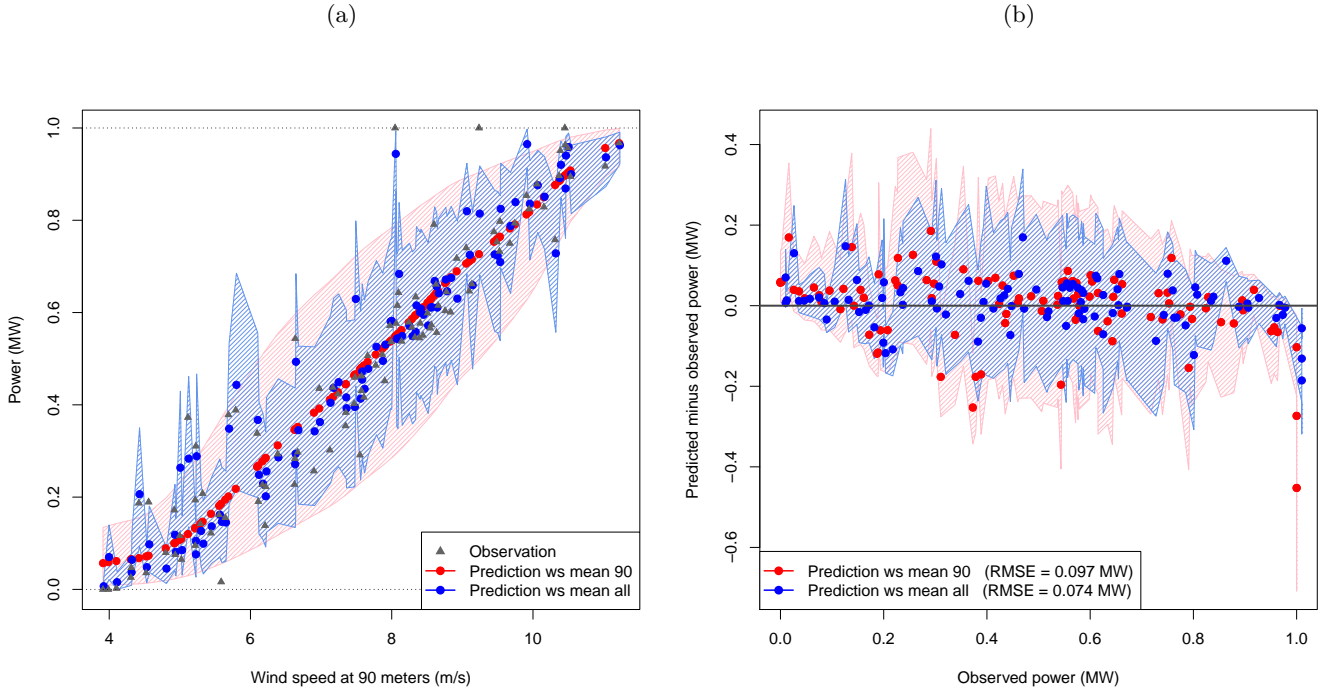


FIG. 8. (a) Predicted power (circles) and the associated 95% prediction intervals (shaded areas) vs. the wind speed mean at 90 meters for a randomly chosen validation set, with predictions obtained with GPM-c model and using either the 90-meter wind speed mean alone (red) or the PCA-transformed entire profile of wind speeds (blue) as inputs. Also shown are the observed power values for this validation set (grey triangles). Note that due to the size of the validation set, only a subset of the validation set is plotted for readability. (b) Wind power residuals (predicted minus observed power) (circles) and the same 95% prediction intervals (shaded areas) as in Fig. 8(a), but shifted by the observed power. The solid grey horizontal line corresponds to perfect prediction.

formation at 90 meters or the REWS, the improvement is substantial, with the median reduction ranging from 22% to 26%. For the 2 larger wind input sets, i.e., those consisting of the entire wind speed profile information, the benefit of adding wind veer, although still statistically significant, is much smaller. The median reductions range from 11% to 16%. This difference in the impact is not surprising: as the amount of wind speed information in the model increases, it captures an increasing amount of information conveyed by wind veer, so the benefit of adding it to the model diminishes. Although very slight, this decrease in benefit is even apparent when the 90-meter wind speed mean is supplemented with the 90-meter wind speed standard deviation (2nd vs. 1st cluster of boxplots in Fig. 10(a)).

The percent reduction magnitudes and patterns in Fig. 10(b) are quite similar to those in Fig. 10(a). For the 3 smaller wind input sets, using air density as a separate input in conjunction with the unadjusted wind inputs (Group 3 in Table II) tends to result in a smaller error than applying the IEC-recommended density adjustment given in Eq. (1). Thus, when wind speed information in the model is limited, allowing the statistical model to determine the nature of the relationship between power and air density and wind speeds is highly advantageous to assuming the relationship in Eq. (2): the RMSE reductions

are positive for all experiments, and their medians range from 22% to 28%. As in the case of wind veer, however, this advantage is much more modest for the two larger wind input sets, with median reductions ranging from 12% to 18%. This is most likely because the model in Eq. (2) is more representative of the reality if the information about the entire wind speed profile—rather than just that at the hub height—is included, so the difference between the two approaches for including air density in the model is smaller.

The plot in Fig. 10(c) shows the improvement resulting from having both wind veer and air density in the model when these are added to the unadjusted wind speed inputs (Group 4 in Table II), relative to using only the air density-adjusted wind speed inputs. The plot shows that when the NN model is used, this combination of inputs tends to have the greatest overall impact of the three groups of additional atmospheric variables considered in the study: the median reductions range from 27% to 31% for the smaller sets and are 21% for the larger sets. For the GPM-c model, as well as the GPM-ns and RF models not shown in these plots, the benefit of using this group of variables is mixed: while the median reductions tend to be similar to those for the other two groups (Figs. 10(a)–(b)), the distribution of these reductions is much wider, so the benefit can be much greater, but also

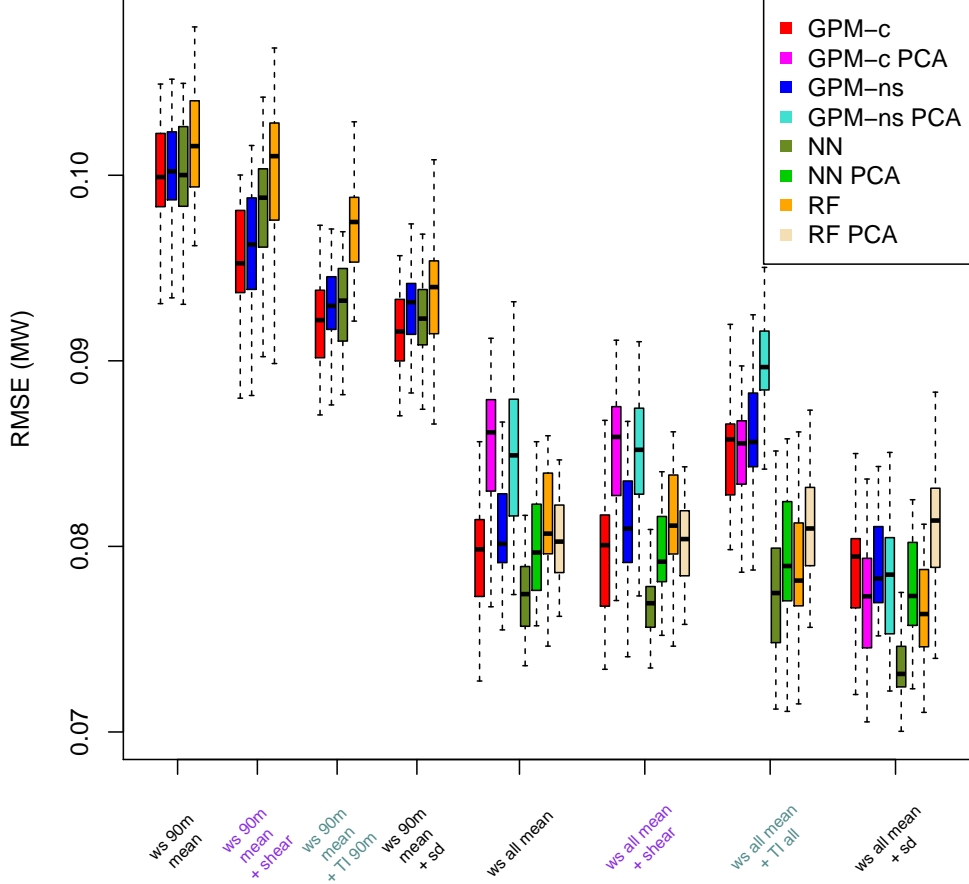


FIG. 9. Boxplots of RMSE values (MW) for 30 experiments using lidar wind speed inputs (sets 1, 2, 4, and 5 in the enumerated list in Section IV B 1, labeled in black on the x -axis) and sets 1 and 4 combined with either shear exponent (labeled in purple on the x -axis) or TI (labeled in teal on the x -axis). When combined with set 1, only the TI at 90 meters was used, whereas TI values at all 12 heights were used when combined with set 4.

much smaller than for the same wind input sets in the other two groups.

C. Caveats

It is important to note several caveats of this study. First, it involved only one location (and one with a complex terrain) and over a period of only 10 weeks. Thus, the above results are a starting point and need to be validated with data that cover longer time periods at other times of year and at other locations with different terrain and turbine types. We are currently acquiring data at one additional location with a much simpler terrain than the one discussed here. We expect that the simpler terrain will alleviate some of the issues we experienced in the course of this study, such as the large discrepancy in inflow conditions measured by the lidar and the nacelle-

mounted anemometer. One of many questions that will be considerably easier to answer as a result of this simplification is which heights in the vertical profile contribute the most valuable wind speed and other atmospheric information for wind power prediction.

Moreover, the analysis was limited to just one turbine directly downwind from the lidar, while ultimately any forecasting tool must predict wind power at multiple turbines at a time, accounting for the wake effects and other spatial and temporal dependencies between them. This would require adjusting the model, e.g., by explicitly including wind direction as an input.

In addition, in this study we explored the relationships between *observed* values of wind speed and other atmospheric variables and power output, as our aim was to better understand the physical processes governing power output. To incorporate these findings and statistical modeling framework into a forecasting tool, how-

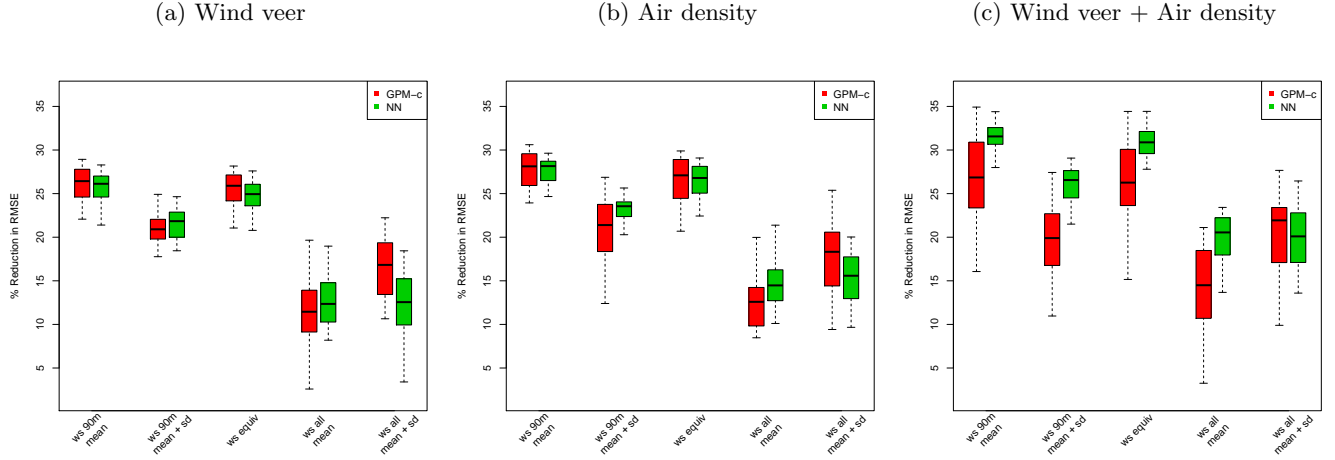


FIG. 10. Boxplots of percent reductions in RMSE (as defined in Section VII B) for the 30 experiments, resulting from using variables in Groups 2–4 of Table II as inputs relative to using the density-adjusted wind speed inputs of Group 1 (shown for the GPM-c and NN models only). Percent reductions resulting from adding (a) **wind veer** to the density-adjusted wind speed inputs (Group 2 vs. 1) (b) **air density** separately to the *unadjusted* wind speed inputs (Group 3 vs. 1), (c) **wind veer and air density** to the *unadjusted* wind speed inputs (Group 4 vs. 1). Note that the scale on the y -axis is the same for all 3 plots.

ever, one would use *forecasts* of the input variables to the power prediction model. The presence of the forecast error in the inputs introduces the need for additional layers of statistical modeling to supplement the framework considered here. For example, a time series approach, such as that taken in Ref. 24 to model the horizontal components of future wind speed, can be combined with our power curve model to convert forecasts of atmospheric conditions into those of wind power.

Another source of error that was not addressed in this study is the measurement error associated with a vertically-profiling lidar. This error can be substantial in complex terrain (up to 10% at the most complex wind farm sites) due to heterogeneous flow across the sample volume⁵³. We are currently refining our approach by explicitly modeling the uncertainty in the wind speed inputs that results either from forecast or measurement error.

Recall that we only focused on the range of wind speed averages that roughly correspond to region II of the standard power curve, i.e., the section between the cut-in and rated wind speeds, where the effects of inflow dominate power production. We did not consider regions I and III since inflow conditions play a much less important role in these regions of the curve. A 10-minute wind speed average value near or below the cut-in speed typically means that for a substantial fraction of the 10-minute period the wind speed was below the cut-in value, resulting in zero power during that portion of the period (since by design the turbine will not generate power below the cut-in wind speed). Thus, forecasting power accurately in region I is primarily a matter of accurately predicting that the wind speed is below the cut-in value.

Similarly, as the wind speed significantly exceeds the rated value, power production becomes increasingly in-

sensitive to the changes in inflow. Instead, the wind turbine control system has a growing impact on power and determines how close the actual power production is to the rated power. However, given the appropriate data, statistical modeling used in this work can be employed to also include the turbine control system settings as inputs to the power curve, so the general framework is still applicable in this region.

Indeed, the variable turbine response can be a significant source of power production variability in all regions of the power curve. For example, the yaw error has been shown to be an important parameter for power prediction⁵⁴. While we did not account for the turbine effects in this work, as noted earlier, we can extend the present framework to include these effects and plan to do so in future work.

In this study we used the RMSE as a metric of performance. While it is a simple and obvious error metric to compute, for some applications it is not as useful as, e.g., a metric that reflects the ability to predict ramp onset, magnitude and duration. Several such metrics have already been proposed and studied (e.g., Ref. 55), and we are currently investigating such alternatives.

Finally, it should be noted that using 10-minute averages may be too coarse to capture certain meteorological phenomena, e.g., fluctuations in air density. We have limited 1-Hz data for the site considered in this study, and we plan to acquire such data at the new site.

VIII. CONCLUSIONS

Despite the caveats discussed above, in this study we have illustrated a general approach for building an improved power curve model. The extensions mentioned in

Section VIIC can be accommodated by tuning or generalizing various aspects of this framework. The results reveal that expanding the set of predictors beyond the density-adjusted free-stream wind speed mean and the use of statistical modeling can lead to substantial improvements in the accuracy of wind power forecasts over the traditionally used power curves.

In particular, wind speed information in the entire vertical profile and wind veer were found to be important predictors of power output. In addition, adding air density separately to the unadjusted wind speeds rather than using the standard IEC-recommended air density adjustment also resulted in more accurate predictions of power. Since the traditional power curve only accommodates the density-adjusted hub-height wind speed mean as an input, expanding the set of predictors requires the use of statistical modeling. However, even if no predictors beyond the hub-height wind speed mean are used, statistical models are still advantageous since they provide estimates of the uncertainty in power predictions.

Results presented here must be validated at other locations with different terrain and turbine types, for longer time periods and other seasons, and we are currently acquiring such data. Analyzing them will likely yield new insights into the importance of various predictors and thus improve the ability to forecast power even further.

ACKNOWLEDGMENTS

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under the Contract DE-AC52-07NA27344 and funded by the LLNL Laboratory Directed Research and Development (LDRD) Grant No. 12-ER-069. This work was supported by the U.S. Department of Energy under Contract DE-AC36-08GO28308 with the National Renewable Energy Laboratory. Funding for the work was provided by the DOE Office of Energy Efficiency and Renewable Energy, Wind and Water Power Technologies Office. The authors are extremely grateful for Infigen Energy's partnership, including data sharing, intellectual input and field campaign assistance. The authors would also like to thank Cary Gellner, Maureen Alai and John Van Fossen (LLNL) for building the solar panel/battery arrays, Jennifer Newman (Univ. of Oklahoma) for assistance in the lidar deployments, and Matthew Simpson, Donald Lucas, Gardar Johannesson (LLNL) and the anonymous reviewers for valuable comments on an earlier version of this manuscript.

Appendix A: Additional details on the statistical models

1. Neural network

A neural network (NN, also known as an artificial neural network, or ANN) is a non-linear regression model

whose mathematical structure mimics the behavior of biological neural networks. In particular, if using a single-layer network, the response Y , such as power output in our case, is modeled as a linear combination of basis functions ϕ_j , $j = 1, \dots, J$, each of which is a non-linear function of predictors \mathbf{x} (such as those discussed in Section IV):

$$Y(\mathbf{x}) = \sum_{j=1}^J w_j \phi_j(\mathbf{x}). \quad (\text{A1})$$

We used the default sigmoid function for the basis functions ϕ_j , whereas the weights w_j and the number of basis functions J were both determined using cross-validation. The library *nnet*⁵⁶ in the R package⁴⁵ was used to fit the NN model. See Refs. 39 and 40, as well as the papers reviewed in Section II, for a more thorough treatment of NNs.

2. Random forest

Like NN, a random forest (RF) is a regression model. This method constructs an ensemble of regression trees, and the individual trees' predictions are averaged to obtain the final prediction. Different trees are constructed by first randomly sampling the training data and then randomly sampling the predictors at each node of the tree. The library *randomForest*⁵⁷ in the R package⁴⁵ was used to fit the RF model. Defaults of various parameters required by the model were used, except for the number of predictors that are randomly sampled at each split, which was tuned to minimize the out-of-bag error. For a more thorough treatment of RFs, see Refs. 40 and 41.

3. Gaussian process model

A general formulation of the stationary Gaussian process model (GPM) is as follows:

$$Y(\mathbf{x}) = \mathbf{g}^T(\mathbf{x})\boldsymbol{\beta} + Z(\mathbf{x}), \quad (\text{A2})$$

where Y is the output, \mathbf{x} is the vector of predictors, \mathbf{g} is a user-defined function, $\boldsymbol{\beta}$ is the vector of the coefficients associated with each predictor, and $Z(\mathbf{x})$ is a stationary error process, modeled as a multivariate Gaussian random variable with 0 mean vector and a user-specified covariance, given by

$$\text{Cov}(\mathbf{x}_i, \mathbf{x}_j) = K(\|\mathbf{x}_i - \mathbf{x}_j\|). \quad (\text{A3})$$

The covariance between any two output values is thus modeled as a user-specified function K of the Euclidean distance between their corresponding predictor vectors \mathbf{x}_i and \mathbf{x}_j , $\|\mathbf{x}_i - \mathbf{x}_j\|$.

The above model specification for the output value $Y(\mathbf{x})$ implies that the joint distribution of the predicted

value $Y(\mathbf{x}_{new})$ at a new set of input values \mathbf{x}_{new} and the observed training set values $y(\mathbf{x}_1), \dots, y(\mathbf{x}_n)$ (in keeping with the standard practice in GPM literature, we denote the predictions and the observed values with the large-case and small-case letters, respectively) is multivariate Gaussian, while the predictive distribution of $Y(\mathbf{x}_{new})$ (i.e., its distribution conditional on the observed response data) is univariate Gaussian. This result is the foundation of GPM prediction inference. See Refs. 42–44 for a detailed treatment of GPM.

The Gaussian random variable is defined on the entire real number line, whereas power output is bounded below by 0 MW and above by the rated power (1 MW in the case of MWT62-1000A wind turbine considered in this work). Consequently, the power values are first transformed using a logit transformation, as follows:

$$Y(\mathbf{x}) = \log \left(\frac{P^*(\mathbf{x})}{1 - P^*(\mathbf{x})} \right) \quad (\text{A4})$$

where $P(\mathbf{x})$ is the power value normalized to the rated power and

$$P^*(\mathbf{x}) = P(\mathbf{x}) \cdot (1 - 2c) + c, \quad (\text{A5})$$

with c a positive constant close to 0. Including c ensures that the logit in Eq. (A4) is defined for normalized power values of 0 and 1 since Eq. (A5) results in $P^*(\mathbf{x})$ of c or $1 - c$ when power $P(\mathbf{x})$ is 0 or 1, respectively. The value of c is chosen so that $Y(\mathbf{x})$ in Eq. (A4) is as close as possible to the Gaussian distribution.

The GPM is then used to obtain the parameters of the predictive Gaussian distribution of $Y(\mathbf{x}_{new})$, which in turn leads to the logit-normal⁵⁸ as the predictive distribution of $P^*(\mathbf{x}_{new})$, i.e., the conditional distribution of $P^*(\mathbf{x}_{new})$ given the observed power values. From Eq. (A5) it follows that

$$P(\mathbf{x}) = \frac{P^*(\mathbf{x}) - c}{1 - 2c}, \quad (\text{A6})$$

so the predictive distribution of $P(\mathbf{x}_{new})$ is a logit-normal shifted by c and scaled by $1 - 2c$. The point prediction of power is then the mean of the predictive distribution of $P(\mathbf{x}_{new})$ since this mean is the minimum mean squared prediction error predictor of $P(\mathbf{x}_{new})$. It is obtained by numerical integration using the *logitnorm* package⁵⁹ in R and applying the transformation on the right-hand side of Eq. (A6). The associated prediction interval is obtained by computing the appropriate quantiles of the predictive Gaussian distribution for $Y(\mathbf{x}_{new})$ and applying the logistic transformation (the inverse of the logit) to them, followed by the transformation in Eq. (A6).

The following covariance function was used for the

TABLE III. Subsets of inputs used to model the mean function and the error process (denoted \mathbf{x}_m and \mathbf{x}_e , respectively) for each of the five input sets enumerated in Section IV when a natural spline of order 3 was used for the function \mathbf{g} . The input labels are as in Section IV B 1.

Input set	\mathbf{x}_m	\mathbf{x}_e
1	ws 90 mean	ws 90 mean
2	ws 90 mean	ws 90 sd
3	ws equiv	ws equiv
4	ws 90 mean	ws all mean
5	ws 90 mean	ws all mean + sd

GPM:

$$\text{Cov}(\mathbf{x}_i, \mathbf{x}_j) = \sigma^2 \left[\exp \left(- \sum_{k=1}^K \frac{|x_{ki} - x_{kj}|^\nu}{d_k} \right) + q\delta_{ij} \right], \quad (\text{A7})$$

where

d_k = range parameter for the k^{th} predictor,

q = nugget and

δ_{ij} = Kroeneker delta

(note that this specification is adapted from Ref. 60). The range (d_k , for each $k = 1, \dots, K$, with K denoting the total number of predictors), the shape (ν), the nugget (q), and the variance (σ^2) parameters were estimated using maximum likelihood estimation. This and other operations needed to implement the GPM were done using routine functions available in the R package⁴⁵.

Finally, we considered two forms of the mean function $\mathbf{g}^T(\mathbf{x})\boldsymbol{\beta}$ in Eq. (A2), resulting from setting \mathbf{g} to a constant or using a natural spline of order 3 (see Refs. 61 and 62) to model it. In case of the former, all the predictors were used to model only the error process. In the latter case, we first separated the predictor set \mathbf{x} into two subsets: one used as inputs to the function \mathbf{g} , denoted with \mathbf{x}_m , and the other used to model the error process Z , denoted with \mathbf{x}_e , as indicated in Table III (note that some predictors were in both sets). We then modeled power output Y as a function of the input vector $\mathbf{x} = (\mathbf{x}_m, \mathbf{x}_e)$, as follows:

$$Y(\mathbf{x}) = Y(\mathbf{x}_m, \mathbf{x}_e) = \mathbf{g}^T(\mathbf{x}_m)\boldsymbol{\beta} + Z(\mathbf{x}_e),$$

with the covariance function as in Eq. (A7).

Appendix B: A Note on the applicability of bootstrap methodology

Bootstrap methodology assumes that the samples are uncorrelated⁴⁸, whereas the RMSEs from the 30 experiments are highly correlated due to the overlap in the data points used across the experiments for both the training and validation sets. However, if for each experiment, we compute the pairwise *difference* in RMSE (between a pair of different input sets and/or models), the correlation among them across the experiments will be minimal under a mild assumption.

If we let R_{i1} and R_{i2} denote the RMSE values for input set/model combination 1 and 2, respectively, and $\Delta_i = R_{i1} - R_{i2}$, $i = 1, \dots, 30$, the covariance of the pairwise differences in two experiments i and j is given by

$$\begin{aligned} \text{Cov}(\Delta_i, \Delta_j) &= \text{Cov}(R_{i1} - R_{i2}, R_{j1} - R_{j2}) \quad (\text{B1}) \\ &= \text{Cov}(R_{i1}, R_{j1}) - \text{Cov}(R_{i1}, R_{j2}) \\ &\quad - \text{Cov}(R_{i2}, R_{j1}) + \text{Cov}(R_{i2}, R_{j2}) \end{aligned}$$

It is reasonable to assume that the main source of correlation between two RMSEs is the overlap in the points used from one experiment to another, i.e., that sharing the same input set and/or model is a much smaller source of this correlation. In that case, we have that

$$\begin{aligned} \text{Cov}(R_{i1}, R_{j1}) &\approx \text{Cov}(R_{i1}, R_{j2}) \\ &\quad \text{and} \quad (\text{B2}) \\ \text{Cov}(R_{i2}, R_{j2}) &\approx \text{Cov}(R_{i2}, R_{j1}). \end{aligned}$$

That is, although the terms on the left-hand sides of Eqs. (B2) are likely higher than those on the right-hand sides because they share the same input set and/or model, we expect the difference between them to be small. It then follows from Eqs. (B1) and (B2) that $\text{Cov}(\Delta_i, \Delta_j) \approx 0$. Since the correlation in the pairwise differences is minimal, we can apply the bootstrap methodology to these differences.

References

- ¹R. Wiser and M. Bolinger, “2012 Wind technologies market report,” Tech. Rep. (U.S. Department of Energy, 2013) <http://emp.lbl.gov/sites/all/files/lbnl-6356e.pdf> (last accessed on 12/15/14).
- ²S. Wharton and J. K. Lundquist, “Atmospheric stability affects wind power collection,” *Environmental Research Letters* **7** 014005 (2012).
- ³J. C. Smith, M. R. Milligan, E. A. DeMeo, and B. Parsons, “Utility wind integration and operating impact state of the art,” *IEEE Transactions on Power Systems* **22**, 900–908 (2007).
- ⁴J. Wang, A. Botterud, R. Bessa, H. Keko, L. Carvalho, D. Issicaba, J. Sumaili, and V. Miranda, “Wind power forecasting uncertainty and unit commitment,” *Applied Energy* **88**, 4014–4023 (2011).
- ⁵A. Papavasiliou and S. S. Oren, “Multi-area stochastic unit commitment for high wind penetration in a transmission constrained network,” *Operations Research* **61**, 578–592 (2013).
- ⁶F. Bingol, J. Mann, and D. Foussekis, “Conically scanning lidar error in complex terrain,” *Meteorologische Zeitschrift* **18**, 189–195 (2009).
- ⁷S. Bradley, “Wind speed errors for lidars and sodars in complex terrain,” *IOP Conference Series: Earth and Environmental Science* **012061** (2008), see <http://stacks.iop.org/1755-1315/1/i=1/a=012061> (last accessed on 11/5/14).
- ⁸S. Bradley, Y. Perrott, P. Behrens, and A. Oldroyd, “Corrections for wind-speed errors from sodar and lidar in complex terrain,” *Boundary-Layer Meteorology* **143**, 37–48 (2012), see <http://dx.doi.org/10.1007/s10546-012-9702-0> (last accessed on 11/8/14).
- ⁹M. Motta, R. Barthelmie, and P. Volund, “The influence of non-logarithmic wind speed profiles on potential power output at Danish offshore sites,” *Wind Energy* **8**, 219–236 (2005).
- ¹⁰J. Sumner and C. Masson, “Influence of atmospheric stability on wind turbine power performance curves,” *Journal of Solar Energy Engineering* **128**, 531–537 (2006).
- ¹¹G. P. van den Berg, “Wind turbine power and sound in relation to atmospheric stability,” *Wind Energy* **11**, 151–169 (2008).
- ¹²I. Antoniou, S. M. Pedersen, and P. B. Enevoldsen, “Wind shear and uncertainties in power curve measurements and wind resources,” *Wind Engineering* **33**, 449–468 (2009).
- ¹³R. Wagner, M. Courtney, J. Gottschall, and P. Lindelow-Marsden, “Accounting for the speed shear in wind turbine power performance measurement,” *Wind Energy* **14**, 993–1004 (2011).
- ¹⁴E. Hedevarang, “Wind turbine power curves incorporating turbulence intensity,” *Wind Energy* **7**, 173–195 (2014).
- ¹⁵B. J. Vanderwende and J. K. Lundquist, “The modification of wind turbine performance by statistically distinct atmospheric regimes,” *Environmental Research Letters* **7** 034035 (2012).
- ¹⁶A. Clifton, L. Kilcher, J. K. Lundquist, and P. Fleming, “Using machine learning to predict wind power output,” *Environmental Research Letters* **8** 024009 (2013).
- ¹⁷A. Clifton and R. Wagner, “Accounting for the effect of turbulence on wind turbine power curves,” in *Proceedings of the Science of Making Torque Conference* (Copenhagen, Denmark, 2014).
- ¹⁸L. Fugon, J. Juban, and G. Kariniotakis, “Data mining for wind power forecasting,” in *Proceedings of the European Wind Energy Conference* (Brussels, Belgium, 2008).
- ¹⁹C. Monteiro, R. Bessa, V. Miranda, A. Botterud, J. Wang, and G. Conzelman, “Wind power forecasting: State-of-the-art 2009,” Tech. Rep. (Argonne National Laboratory, ANL/DIS-10-1, 2009).
- ²⁰P. M. Fonte and J. C. Quadrado, “ANN approach to WECS power forecast,” in *Proceedings of the 10th IEEE Conference on Emerging Technologies and Factory Automation* (Catania, Italy, 2005).
- ²¹K. Rohrig and B. Lange, “Application of power prediction tools for power system operations,” in *Proceedings of the IEEE Power Engineering Society General Meeting* (Montreal, Canada, 2006).
- ²²M. C. Alexiadis, P. S. Dokopoulos, and H. S. Sahsamanoglou, “Wind speed and power forecasting based on spatial correlation models,” *IEEE Transactions on Energy Conversion* **14**, 836–842 (1999).
- ²³S. Li, D. C. Wunsch, E. A. O’Hair, and M. G. Giesselmann, “Using neural networks to estimate wind turbine power generation,” *IEEE Transactions on Energy Conversion* **16**, 276–282 (2001).
- ²⁴J. Jeon and J. W. Taylor, “Using conditional kernel density estimation for wind power density forecasting,” *Journal of the American Statistical Association* **107**, 66–79 (2012).
- ²⁵Y. V. Makarov, D. L. Hawkins, E. Leuze, and J. Vidov, “California ISO wind generation forecasting service design and experience,” in *Proceedings of the AWEA Conference* (Portland, OR, 2002).
- ²⁶M. Milligan, M. Schwartz, and Y. H. Wan, “Statistical power forecasting models: Results for U.S. wind farms,” Tech. Rep. (National Renewable Energy Laboratory, CP-500-33956, 2003).
- ²⁷M. J. Duran, D. Cros, and J. Riquelme, “Short-term wind power forecast based on ARX models,” *Journal of Energy Engineering* **133**, 172–180 (2007).
- ²⁸P. Pinson, “Wind energy: Forecasting challenges for its operational management,” *Statistical Science* **28**, 564–585 (2013).
- ²⁹M. Yoder, A. S. Hering, W. C. Navidi, and K. Larson, “Short-term forecasting of categorical changes in wind power with Markov chain models,” *Wind Energy* **17**, 1425–1439 (2013).
- ³⁰“Mitsubishi wind turbine generator specifications,” see <https://www.mhi-global.com/products/pdf/mwt1000a.pdf> (last accessed on 11/8/14).
- ³¹R. Krishnamurthy and M. Boquet, “Case studies of WINDCUBE measurement uncertainty for complex terrain using FLOW complexity recognition (FCR),” Poster presentation at the EWEA conference in Barcelona, Spain, 2014.
- ³²S. Wharton, G. Qualley, J. Newman, and W. Miller, “2013 LIDAR campaign at an Altamont Pass wind farm: An examination of hill “speedup” flows,” Tech. Rep. (Lawrence Livermore National Laboratory, LLNL-TR-644242, 2013).

- ³³“International Electrotechnical Commission (IEC) 61400-12-1:2005(e): Wind turbines - Part 12-1: Power performance measurements of electricity producing wind turbines,” (2005).
- ³⁴“International Electrotechnical Commission (IEC) International Standard 61400-12-1: Wind turbine generator systems - Wind turbine power performance testing,” (1998).
- ³⁵R. Wagner, I. Antoniou, S. M. Pedersen, M. S. Courtney, and H. E. Jorgensen, “The influence of the wind speed profile on wind turbine performance measurements,” *Wind Energy* **12**, 348–362 (2009).
- ³⁶N. Kelley, M. Shirazi, D. Jager, S. Wilde, J. Adams, and M. Buhl, “Lamar low-level jet project interim report,” Tech. Rep. (National Renewable Energy Laboratory, NREL/TP-500-34593, 2004).
- ³⁷I. T. Jolliffe, *Principal Component Analysis* (Springer, 2002).
- ³⁸J. R. Jammalamadaka and A. SenGupta, *Topics in circular statistics* (World Scientific Press, 2001) pp. 176–179.
- ³⁹C. Bishop, *Pattern Recognition and Machine Learning* (Springer, 2007) pp. 225–281.
- ⁴⁰T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning* (Springer, 2009) pp. 587–602.
- ⁴¹L. Breiman, “Random forests,” *Machine Learning* **45**, 5–32 (2001).
- ⁴²J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn, “Design and analysis of computer experiments,” *Statistical Science* **4**, 409–423 (1989).
- ⁴³C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning* (The MIT Press, 2005).
- ⁴⁴T. J. Santner, B. J. Williams, and W. I. Notz, *The Design and Analysis of Computer Experiments* (Springer, 2003).
- ⁴⁵R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria (2013), see <http://www.R-project.org> (last accessed on 11/8/14).
- ⁴⁶R. B. Gramacy and H. K. H. Lee, “Bayesian treed Gaussian process models with an application to computer modeling,” *Journal of the American Statistical Association* **103**, 1119–1130 (2008).
- ⁴⁷R. B. Gramacy, “tgp: An R package for Bayesian nonstationary, semiparametric nonlinear regression and design by treed Gaussian process models,” *Journal of Statistical Software* **19**, 9 (2007).
- ⁴⁸B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap* (Chapman & Hall/CRC, 1998) pp. 184–188.
- ⁴⁹R. Tibshirani, “Package ‘bootstrap’: Functions for the book ‘An Introduction to the Bootstrap’,” see <http://cran.r-project.org/web/packages/bootstrap/bootstrap.pdf> (last accessed on 11/8/14).
- ⁵⁰Y. Benjamini and Y. Hochberg, “Controlling the false discovery rate: a practical and powerful approach to multiple testing,” *Journal of Royal Statistical Society, Series B* **57**, 289–300 (1995).
- ⁵¹S. Wharton and J. K. Lundquist, “Assessing atmospheric stability and its impacts on rotor-disk wind characteristics at an onshore wind farm,” *Wind Energy* **15**, 525–546 (2012).
- ⁵²R. Nierenberg, “Free-flow variability on the Jess and Souza Ranches, Altamont Pass,” Tech. Rep. (U.S. Department of Energy/Solar Energy Research Institute, STR-217-3404, 1989).
- ⁵³F. Bingol, J. Mann, and D. Foussekis, “Modeling conically scanning lidar error in complex terrain with WAsP Engineering,” Tech. Rep. (Riso DTU, Riso-R-1664(EN), 2008).
- ⁵⁴K. A. Kragh, P. A. Fleming, and A. Scholbrock, “Increased power capture by rotor speed-dependent yaw control of wind turbines,” *Journal of Solar Energy Engineering* **135** 031018 (2013).
- ⁵⁵C. Ferreira, J. Gama, L. Matias, A. Botterud, and J. Wang, “Survey and wind power ramp forecasting,” Tech. Rep. (Argonne National Laboratory, ANL/DIS-10-13, 2010).
- ⁵⁶W. N. Venables and B. D. Ripley, *Modern Applied Statistics with S*, 4th ed. (Springer, New York, 2002) ISBN 0-387-95457-0.
- ⁵⁷A. Liaw and M. Wiener, “Classification and regression by randomforest,” *R News* **2**, 18–22 (2002), see <http://CRAN.R-project.org/doc/Rnews/> (last accessed on 11/8/14).
- ⁵⁸R. Mead, “A generalised logit-normal distribution,” *Biometrics* **21**, 721–732 (1965).
- ⁵⁹T. Wutzler, “Package ‘logitnorm’: Functions for the logitnormal distribution,” see <http://cran.r-project.org/web/packages/logitnorm/logitnorm.pdf> (last accessed on 11/8/14).
- ⁶⁰R. B. Gramacy and H. K. H. Lee, “Adaptive design and analysis of supercomputer experiments,” *Technometrics* **51**, 130–145 (2009).
- ⁶¹T. Hastie, *Generalized Additive Models* (Chapman and Hall/CRC, 1990) pp. 27–28.
- ⁶²J. M. Chambers and T. J. Hastie, *Statistical Models in S* (Wadsworth and Brooks/Cole, 1992) pp. 270–273.